Automatic Prediction of Spirometry Readings from Cough and Wheeze for Monitoring of Asthma Severity

Achuth Rao MV, Kausthubha NK, Shivani Yadav, Dipanjan Gope, Uma Maheswari Krishnaswamy, Prasanta Kumar Ghosh

Abstract—We consider the task of automatically predicting spirometry readings from cough and wheeze audio signals for asthma severity monitoring. Spirometry is a pulmonary function test used to measure forced expiratory volume in one second (FEV1) and forced vital capacity (FVC) when a subject exhales in the spirometry sensor after taking a deep breath. FEV1%, FVC% and their ratio are typically used to determine the asthma severity. Accurate prediction of these spirometry readings from cough and wheeze could help patients to non-invasively monitor their asthma severity in the absence of spirometry. We use statistical spectrum description (SSD) as the cue from cough and wheeze signal to predict the spirometry readings using support vector regression (SVR). We perform experiments with cough and wheeze recordings from 16 healthy persons and 12 patients. We find that the coughs are better predictor of spirometry readings compared to the wheeze signal. FEV1%, FVC% and their ratio are predicted with root mean squared error of 11.06%, 10.3% and 0.08 respectively. We also perform a three class asthma severity level classification with predicted FEV1% and obtain an accuracy of 77.77%.

I. INTRODUCTION

Asthma is a chronic inflammatory disease of the airways caused by the combination of genetic and environmental factors like air pollution or allergens [15]. World Health Organization (WHO) estimates that 235 million people currently suffer from asthma, with 250k annual deaths attributed to the disease [19]. Asthma severity is clinically categorized into four classes – intermittent, mild persistent, moderate persistent, severe persistent – according to the frequency of symptoms, forced expiratory volume in one second (FEV1), forced vital capacity (FVC) and FEV1 to FVC ratio (FEV1/FVC). The symptoms and the threshold values for different levels of severity are given in [7], [26]. The severe persistent asthma is life threatening.

Spirometry is the most common of the pulmonary function tests and specifically measures the FEV1 and FVC. There are reference values of the FEV1 (FEV1_ref) and FVC (FVC_ref) for each patient depending on his/her age, gender, height and weight. The FEV1% and FVC% denote the ratio (in percentage) between the value measured by the spirometry and the corresponding reference value. FEV1%, FVC% and the ratio of FEV1 & FEV (FEV1/FVC) are indicators of the severity of asthma. For spirometry readings, the patients are asked to take a deep breath to the best of their capacity, and then exhale into the sensor as fast and long as possible, preferably at least for 6 seconds. It is believed to be the single best test for asthma [18]. But the maneuver primarily depends on patient’s cooperation and effort, causing the readings to vary depending on how meticulously a patient does the inhalation and exhalation in the suggested manner. So it becomes difficult to obtain spirometry readings for children and elderly people [10]. It is also often required for the asthma patients to monitor their asthma level at home [25]. However, the spirometry is expensive and not a portable device. Thus, the peak flow meter (PFM) [1] is often used as a substitute which measures how well the lungs push out air. But it is known that a PFM is less accurate than spirometry [8]. A PFM can only measure the air flow through the major airways of patient’s lungs. These major airways are those from where the strength of exhalation comes. However, minor airways in one’s lungs could be affected by asthma in a manner similar to the major airways resulting in minor airways to swell causing typical asthma symptoms. But a PFM fails to measure the strength of those airways. Thus, it would be useful to have a cheap and portable device that could measure FEV1 and FVC as good as spirometry.

The cough is produced by closing the glottis till the pressure builds up below the glottis followed by a sudden release of pressure once the glottis opens. Wheeze, on the other hand, is a continuous flow of air from lungs to the mouth. For both cough and wheeze, the air volume that flows from lungs to the mouth is modulated by the obstruction in the airways caused by asthma. So we hypothesize that the severity of asthma could be predicted from the wheeze/cough sound. Cough and wheeze could be easily recorded by a microphone, often available in the smart phones. Thus, predicting spirometry values from cough and wheeze would be non-invasive. It would also be comfortable unlike using spirometry irrespective of the age and medical condition of the patient.

There are a number of works in the literature for classifying a subject into asthmatic or healthy person based on his/her
cough and wheeze. For example, Wisniewski et al. used tonal index to detect pulmonary wheezes for asthma monitoring [24]. Similarly, Holmes et al. automatic identification of inhalations in asthma inhaler recordings [14]. Akram et al [3] proposed a segmentation scheme of respiratory sounds for the detection of wheezes for asthma detection. Study carried out by Bentur et al [4] shows how wheeze monitoring provides quantitative information that correlates well with asthma activity of children. On the other hand, several algorithms have been proposed to identify coughs [13], [16] for asthma detection. Batra et al. [2] explored features such as harmonic to noise ratio (HNR), jitter and shimmer in the sustained vowel phonation for identifying asthma patients. There are several works that classify asthma using respiratory sound based on pitch [6], [20], dominant frequency range [11], [12] and duration of the breath [22]. To the best of our knowledge, there is no work reported on predicting spirometry readings from cough and wheeze for asthma level monitoring.

We, in this work, have explored the task of predicting spirometry readings based on statistical spectrum descriptor (SSD) [9] from cough and wheeze signal. We have used support vector regression [23] to predict the spirometry readings from the SSDs of the wheeze and cough sound. Experiments are performed in a leave-one-subject-out setup with cough and wheeze recordings from 16 healthy subjects and 12 asthmatic patients. We find that, on average, FEV1%, FVC% and FEV1/FVC are predicted with a root mean squared error of 11.6%, 10.3%, and 0.08 respectively. We also perform a three-class asthma severity classification using the predicted FEV1% and obtain an accuracy of 77.77%, which turns out to be ~16% (absolute) higher than the baseline scheme. We also perform a feature selection to investigate the subset of features that provides maximal information of severity of asthma. We begin with the description of the dataset.

II. DATASET

The recordings used in this study were obtained from a total 28 subjects comprising 16 healthy subjects (10 male and 6 female) and 12 asthmatic patients (7 male and 5 female) recruited from St. John’s National Academy of Health Sciences, Bangalore. The healthy subjects were middle aged with an age range of 19-37 years with an average age of 26 years. The age range of the patients was 19-75 years with an average age of 41 years. On doctor’s suggestion, the patients go through the standard spirometry [17] to measure the FEV1 and FVC. Prior approval for recording was obtained from hospital ethics committee and consent for recording was taken from each subject. Following spirometry test, subjects are asked to cough and wheeze for at least five times, which were recorded at a sampling rate of 48kHz and 16-bit using the ZOOM H6 handy recorder. Sufficient break was given between the spirometry test and the cough/wheeze recording to ensure that the patient is comfortable during recording. The start and end of each wheeze/cough are manually marked. Among 12 patients, six patients have the recording both before and after the bronchodilator. The range of FEV1%, FVC% and FEV1_FVC of all subjects were 28-100%, 35-100% and 62-100% with their average values of 70%, 68%, and 87% respectively. Sample wheeze and cough signals along with their spectrograms are shown in the Fig. 1. The inhalation and exhalation in wheeze sound can be clearly seen. The spectrogram also shows the time-varying spectral content. The time-frequency characteristics of cough sound appears to be different from that of wheeze sound. From the entire recording, on average, we obtain 6(±5) wheeze and 6(±5) cough recordings per subject.

III. PROPOSED APPROACH FOR PREDICTING SPIROMETRY READINGS FROM COUGH AND WHEEZE

The block diagram of the proposed approach is shown in Fig. 2. For a given recording from a subject, there are multiple instances of the cough/wheeze sounds and one set of the spirometry readings. Using a regression model on the acoustic features of cough/wheeze, we predict the spirometry reading for each instance of cough/wheeze and combine the predicted values in the end to get the final estimate of the spirometry reading for the subject. Thresholds are applied on the spirometry reading, FEV1%, to perform the asthma severity classification. The each block of Fig. 2 are explained in details below.
FVC (target variables). SVR \([23]\) is an
FVC are unitless. We also use
FVC\%) for different
We use the root mean squared error
\[ \text{RMSE} = \sqrt{\frac{1}{L} \sum_{l=1}^{L} (x_l - \hat{x}_l)^2} \]
B. Regression model
performance compared to that using SSDs.
\[ \text{RMSE} \] lie within an
\[ \text{standard deviation (SD)} \]
A. Feature extraction
We explore the widely used feature in speech, namely, Mel-frequency cepstral coefficients (MFCCs). MFCC is ex-
ttracted by first extracting logarithm of energies in sub-bands
placed uniformly on the mel-scale and then computing the
discrete cosine transform (DCT). The DCT provides a low-
dimensional representation compared to the number of sub-
bands. Mel-scale reflects the nonlinear frequency sensitivity
of the human auditory system \([9]\). We extract MFCC for short
overlapping segments resulting in a sequence of MFCCs for
each cough/wheeze recording. For regression, the sequence
of MFCCs are converted to a single vector by computing an
average value for each element in the MFCC vectors in the
sequence to obtain statistical spectrum descriptor (SSD) of a
cough/wheeze instance. In addition to taking average, we have
experimented with other statistics of the MFCCs including
variance, median, however, there was no improvement in the
performance compared to that using SSDs.

B. Regression model
Support vector regression (SVR) is used as the regression
model. We have explored nonlinear \(\epsilon\)-SVR to approximate
the function between the SSDs and FEV1, FEV\%, FVC,
FVC\% and FEV1_FVC (target variables). SVR \([23]\) is an
application of SVM to find the mapping function between
input and output. We use \(\epsilon\)-SVR, which tries to find the optimal
regression hyperplane so that most of the training samples
lie within an \(\epsilon\)-margin around this hyperplane. Non-linear
regression is done in an efficient way by applying the Kernel
function, i.e., to replace the inner product in the solution by a
non-linear kernel function. We used the radial basis function
as kernel for regression and used LIBSVM \([5]\) toolkit for SVR
implementation.

C. Final prediction and asthma severity classification
The SVR is trained with the SSDs and spirometry readings
from the training set. We obtain the predicted values of the
spirometry readings for each cough/wheeze instance in the test
set for a subject. The final spirometry reading is computed
by taking the median of these predicted values across all
instances. Among the spirometry readings, FEV1\% is used
to perform asthma severity level by using the predefined
thresholds \([26]\) for asthma severity levels.

IV. EXPERIMENTS AND RESULTS
A. Experimental Setup
Each cough/wheeze realization is windowed with 25ms win-
dow and 10ms shift to compute 13-dimensional MFCC, which
is computed by using sub-bands placed uniformly on mel-
scale in the range of 100Hz-3700Hz. Every coefficient in the
MFCC vector sequence is averaged across all frames in each
cough/wheeze sound to obtain a 13-dimensional SSD vector
as the acoustic representation. The target spirometry readings
(FEV1, FEV1\%, FVC, FVC\%, FEV1_FVC\%) for different
cough/wheeze sound from the same subject are identical since
only one set of spirometry reading is available for a subject.

For SVR, a leave-one-subject-out cross validation is used to
examine the robustness of the predictive model to the variation
due to speaker characteristics. The hyper parameters \(\epsilon, C\) and
\(\sigma\) are optimized by grid search to maximize the performance
on the training data. The value of \(\epsilon, C\) and \(\sigma\) lie within a range
of \(10^{-6} - 0.1, 0.01 - 1000 \) and \(10^{-4} - 1\) respectively.

Evaluation Metrics: We use the root mean squared error
(RMSE) between the ground truth spirometry readings and
the predicted one. Across all test cases in the leave-one-
subject-out setup. We also report standard deviation (SD) of
the squared errors which captures the variation of the squared
errors around the RMSE. Suppose there are \(L\) test cases,
where \(x_l, 1 \leq l \leq L\) and \(\hat{x}_l, 1 \leq l \leq L\) denote
the original and predicted spirometry values. Then the RMSE
and SD are defined as follows:
\[ \text{RMSE} = \sqrt{\frac{1}{L} \sum_{l=1}^{L} (x_l - \hat{x}_l)^2} \]
and
\[ \text{SD} = \sqrt{\frac{1}{L} \sum_{l=1}^{L} ((x_l - \hat{x}_l)^2 - \text{RMSE}^2)}^2 \].
The units of RMSE and SD for FEV1 and FVC are liters, while those
FEV1\%, FVC\% and FEV1_FVC\% are unitless. We also use
the classification accuracy as a metric for the asthma severity
classification. For asthma severity classification, we consider
three classes by using two thresholds of 0.8 and 0.6 on

Fig. 3. Illustration of the selected features for predicting different spirometry readings using cough and wheeze. In each subplot, x-axis denotes the
indices of the test subject in leave-one-subject-out setup. Y-axis denotes the SSD index. A white box for a particular SSD and test subject indicates that the corresponding
SSD is selected for the respective test subject.

RMSE = \sqrt{\frac{1}{L} \sum_{l=1}^{L} (x_l - \hat{x}_l)^2} \]
and
\[ \text{SD} = \sqrt{\frac{1}{L} \sum_{l=1}^{L} ((x_l - \hat{x}_l)^2 - \text{RMSE}^2)}^2 \]. The units of
RMSE and SD for FEV1 and FVC are liters, while those
FEV1\%, FVC\% and FEV1_FVC\% are unitless. We also use
the classification accuracy as a metric for the asthma severity
classification. For asthma severity classification, we consider
three classes by using two thresholds of 0.8 and 0.6 on

IV. EXPERIMENTS AND RESULTS
A. Experimental Setup
Each cough/wheeze realization is windowed with 25ms win-
dow and 10ms shift to compute 13-dimensional MFCC, which
is computed by using sub-bands placed uniformly on mel-
scale in the range of 100Hz-3700Hz. Every coefficient in the
MFCC vector sequence is averaged across all frames in each
cough/wheeze sound to obtain a 13-dimensional SSD vector
as the acoustic representation. The target spirometry readings
(FEV1, FEV1\%, FVC, FVC\%, FEV1_FVC\%) for different
cough/wheeze sound from the same subject are identical since
only one set of spirometry reading is available for a subject.

For SVR, a leave-one-subject-out cross validation is used to
examine the robustness of the predictive model to the variation
due to speaker characteristics. The hyper parameters \(\epsilon, C\) and
\(\sigma\) are optimized by grid search to maximize the performance
on the training data. The value of \(\epsilon, C\) and \(\sigma\) lie within a range
of \(10^{-6} - 0.1, 0.01 - 1000 \) and \(10^{-4} - 1\) respectively.

Evaluation Metrics: We use the root mean squared error
(RMSE) between the ground truth spirometry readings and
the predicted one. Across all test cases in the leave-one-
subject-out setup. We also report standard deviation (SD) of
the squared errors which captures the variation of the squared
errors around the RMSE. Suppose there are \(L\) test cases,
where \(x_l, 1 \leq l \leq L\) and \(\hat{x}_l, 1 \leq l \leq L\) denote
the original and predicted spirometry values. Then the RMSE
and SD are defined as follows:
\[ \text{RMSE} = \sqrt{\frac{1}{L} \sum_{l=1}^{L} (x_l - \hat{x}_l)^2} \]
and
\[ \text{SD} = \sqrt{\frac{1}{L} \sum_{l=1}^{L} ((x_l - \hat{x}_l)^2 - \text{RMSE}^2)}^2 \]. The units of\nRMSE and SD for FEV1 and FVC are liters, while those
FEV1\%, FVC\% and FEV1_FVC\% are unitless. We also use
the classification accuracy as a metric for the asthma severity
classification. For asthma severity classification, we consider
three classes by using two thresholds of 0.8 and 0.6 on
We investigate the predictive power of the FEV1% [26]. Although there are four clinically defined asthma severity levels – intermittent (FEV1% > 0.8), mild persistent (FEV1% > 0.8), moderate persistent (0.6 < FEV1% < 0.8), severe persistent (FEV1% < 0.6) – the differences between intermittent and mild persistent lie in the symptoms only [26]. Thus we merge these two levels to result in three asthma severity classes. For predicting spirometry readings we consider a vanilla baseline where the average in three asthma severity classes. For predicting spirometry readings we consider a vanilla baseline where the average of the all readings in the training set is used. Similarly, for asthma severity classification, we classify each test case with ‘moderate persistent’ asthma since the maximum number data belongs to this class.

**Feature selection:** We investigate the predictive power of each coefficient of the SSD. For this purpose, we have used a forward feature selection algorithm following the work by Abhay et al. [21]. For the feature selection, a three fold cross-validation is performed within the training set where only one subject’s data is used as the test set in the leave-one-subject-out setup. The features, thus selected, are used for the unseen test data. The forward feature selection is performed separately for each target variables, i.e., FEV1, FEV1%, FVC, FVC%, FEV1_FVC.

### Table I

<table>
<thead>
<tr>
<th>Spirometry readings</th>
<th>FEV1</th>
<th>FEV1%</th>
<th>FVC</th>
<th>FVC%</th>
<th>FEV1_FVC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0.77 (±0.69)</td>
<td>15.24 (±4.2)</td>
<td>0.85 (±0.98)</td>
<td>13.80 (±3.4)</td>
<td>0.08 (±1.2)</td>
</tr>
<tr>
<td>w/feature selection</td>
<td>wheeze</td>
<td>0.70 (0.65)</td>
<td>13 (1.10)</td>
<td>0.77 (0.98)</td>
<td>13 (3.25)</td>
</tr>
<tr>
<td></td>
<td>cough</td>
<td>0.48 (0.24)</td>
<td>12.1 (1.77)</td>
<td>0.57 (0.46)</td>
<td>12.4 (2.57)</td>
</tr>
<tr>
<td>w/feature selection</td>
<td>wheeze</td>
<td>0.66 (0.74)</td>
<td>12 (4.41)</td>
<td>0.74 (1.04)</td>
<td>12 (2.86)</td>
</tr>
<tr>
<td></td>
<td>cough</td>
<td>0.48 (0.31)</td>
<td>11.6 (1.64)</td>
<td>0.57 (0.53)</td>
<td>10.3 (1.99)</td>
</tr>
</tbody>
</table>

**B. Results and discussion**

The RMSE of different spirometry readings predicted from wheeze and cough using baseline technique, SSDs and selected SSDs are shown in Table 1. The entries in the table indicate the RMSE across all test cases in the leave-one-subject-out setup. It is clear that the RMSE obtained using SSD reduces compared to that using the baseline scheme for all the spirometry readings except for FEV1_FVC where the RMSE using SSDs of wheeze is identical to that using baseline scheme. This suggests that the spectral characteristics captured by SSD are indicative of the variation of the spirometry readings due to different asthma severity levels. It is also interesting to observe that the RMSE using cough is consistently lower than that using wheeze indicating the cough to be a better predictor of the spirometry readings. In fact, the SD also reduces for using cough compared to wheeze.

Fig. 3 summarizes the features selected using forward feature selection algorithm corresponding to different spirometry readings. It is interesting to note that, on average, the number of features selected for cough based prediction is higher than that for wheeze based prediction. In fact, a few SSDs are consistently selected for all test subjects. For example, 2nd SSD is selected for all test subjects as well as for all spirometry readings. 2nd SSD is computed from 2nd MFCC which captures the spectral tilt in the range of 100-3700Hz. This indicates that the spectral tilt in the cough signal could be a good indicator of the asthma severity. In the case of wheeze based prediction, no SSD gets consistently selected for all test subjects. However, the 1st SSD has been selected for most of the test subjects for predicting FEV1%, FVC% and FEV1_FVC. Similarly, the 4th SSD turns out to be the maximally selected feature for FEV1 and FVC. From Table 1, it is clear that the RMSE drops with selected feature compared to those without feature selection for most of the spirometry readings. This suggests that few information bearing SSDs could predict the spirometry readings with lower RMSE than that using all SSDs. It is also clear from Table 1 that the RMSE obtained by using a cough signal is consistently lower than that using all SSDs.

**Feature selection:** We investigate the predictive power of each coefficient of the SSD. For this purpose, we have used a forward feature selection algorithm following the work by Abhay et al. [21]. For the feature selection, a three fold cross-validation is performed within the training set where only one subject’s data is used as the test set in the leave-one-subject-out setup. The features, thus selected, are used for the unseen test data. The forward feature selection is performed separately for each target variables, i.e., FEV1, FEV1%, FVC, FVC%, FEV1_FVC.

### Table II

**Asthma severity classification accuracy (in %), the dimension of the ranked SSDs is two and eleven for cough and wheeze respectively.**

<table>
<thead>
<tr>
<th>Feature selection</th>
<th>wheeze</th>
<th>cough</th>
</tr>
</thead>
<tbody>
<tr>
<td>13-dim SSDs</td>
<td>67.85</td>
<td>74.07</td>
</tr>
<tr>
<td>selected SSDs</td>
<td>67.85</td>
<td>74.07</td>
</tr>
<tr>
<td>ranked SSDs</td>
<td>57.14</td>
<td>77.77</td>
</tr>
<tr>
<td>baseline</td>
<td>61.76</td>
<td></td>
</tr>
</tbody>
</table>
those using wheeze signal even with selected features. This reaffirms that cough is a better predictor of asthma severity compared to wheeze.

The feature selection is done separately for each test subject using the respective training set. This results in variations in the selected SSDs across different training sets in the leave-one-subject-out setup. This could also lead to overtraining causing poor performance on the test subject. For this purpose, we rank order the SSDs in the decreasing order of their accuracies as selected features across different training sets. We then use top $K(1 \leq K \leq 13)$ SSDs from this ranked list of SSDs and use this as a fixed set of features for all test subjects. The RMSE vs. $K$ for different spirometry readings from cough and wheeze signals is shown in Fig. 4. From the figure, it is clear that the minimum RMSE for predicting FEV1, FEV1%, FVC, FVC%, and FEV1_FVC are 0.63, 8.71, 0.64, 8.73, 0.09 from wheeze and 0.45, 11.53, 0.55, 12.1, 0.08 from cough respectively. These RMSEs are lower for most of the spirometry readings compared to those in Table 1 with selected features. This could imply that the set of SSDs corresponding to the minimum RMSE in Fig. 4 could be more robust to variation in test subjects compared to SSDs selected from the training set separately for each test subject.

Using the predicted FEV1%, the asthma severity classification accuracy in the leave-one-subject-out setup is given in Table 2. Classification accuracies are reported using SSDs, selected SSDs, top few ranked SSDs (two for cough and ten for wheeze). It is clear that the classification accuracies using both cough and wheeze are better than baseline; however, accuracy using cough is better than that using wheeze for using selected and ranked SSDs. This is mainly because the RMSE of FEV1% is lower using cough than using wheeze.

V. CONCLUSIONS AND FUTURE WORK

We present a technique for predicting spirometry readings and asthma severity classification based on cough and wheeze sound using SSD as the acoustic feature and SVR as the regression model. The proposed approach predicts FEV1%, FVC%, and FEV1_FVC with RMSE of 11.6%, 10.3%, and 0.08 respectively. The three-class asthma severity classification using the predicted spirometry readings results in a classification accuracy of 77.77%. Further improvement on the RMSE can be made by considering the temporal evolution of MFCC in each cough/wheeze signal. The presented technique for predicting spirometry readings could be integrated with automatic cough and wheeze detector to automatically predict and monitor the asthma severity from a subject’s voice. This is part of our future work.

REFERENCES


